

User-Centered Evaluations with Older Adults: Testing the Usability of a Mobile Health System for Heart Failure Self-Management

Victor P. Cornet, MS¹, Carly N. Daley, MS¹, Preethi Srinivas, PhD², Richard J. Holden, PhD^{1,2*}

¹Indiana University School of Informatics and Computing - Indianapolis

²Indiana University Center for Aging Research, Regenstrief Institute, Inc.

*Corresponding author: rjholden@iupui.edu

Many older adults living with heart failure struggle to follow recommended self-management routines. To help older adults with heart failure more effectively and efficiently self-manage their disease, we developed Engage, a mobile health application promoting the performance, logging, and sharing of routine self-management behaviors. This paper reports on the usability evaluation of the Engage system with 15 older adults with heart failure and informal caregivers. In two phases, participants used Engage during a task-based usability test (n=5) and a scenario-based usability test (n=10). Usability and performance data were assessed through video-recorded observation and the administration of the system usability scale (SUS) and NASA Task Load Index (TLX). We found that task-based testing was useful in quickly identifying problems within our application, but scenario-based testing elicited more valuable feedback from older adults. A comparison of the different evaluation methods used and the discussion of the challenges encountered provide multiple implications for the practice of usability testing of mobile health products with older adults.

INTRODUCTION

Millions of older adults with heart failure are expected to follow complex self-management recommendations and could benefit from a tool to routinize and simplify self-management. Mobile technologies can be leveraged for this purpose because of their convenience and efficiency, but only if they have been thoroughly evaluated for usability (Fisk, Rogers, Charness, Czaja, & Sharit, 2009). The objective of this study was **to apply a series of usability evaluation methods to test Engage, a mobile health (mHealth) system for older adults self-managing heart failure**. We identified both specific usability issues during testing and methodological implications revealed by our experiences with each method.

Heart Failure Self-Management

Heart failure is a complex clinical diagnosis affecting over 5.8 million people in the U.S. (Roger, 2013). Most patients with heart failure are aged ≥ 65 years and heart failure prevalence has risen as the population has aged (Chaudhry et al., 2013). Patients with heart failure are expected to perform a complex self-management regimen, consisting of medications, a sodium-restricted diet, fluid intake restriction, daily recording of weight and vitals, exercise, and continual self-monitoring for symptoms (Lainscak et al., 2011). Such tasks sometimes involve informal caregivers, such as domestic partners and children (Mickelson & Holden, 2013). There is ample evidence that older adults do not adhere to the above self-management recommendations (van der Wal & Jaarsma, 2008; van der Wal, Jaarsma, & van Veldhuisen, 2005). Reported reasons for non-adherence include personal factors such as lacking knowledge and motivation (van der Wal et al., 2006; Wu, Moser, Chung, & Lennie, 2008) as well as system barriers such as task burden, lack of tools, and inadequate resources (Granger, Sandelowski, Tahshjain, Swedberg, & Ekman, 2009; R.J. Holden et al., 2015; Mickelson, Willis, & Holden, 2015). Further, it has been shown that those successful in self-managing heart failure are ones

who have established self-management habits and whose performance is therefore more routinized (Mickelson & Holden, 2017; Riegel, Dickson, & Topaz, 2013).

mHealth for Geriatric Heart Failure Self-Management

mHealth technologies could help routinize and alleviate the burden of heart failure self-management. However, mHealth products suffer from high rates of system discontinuation (Eysenbach, 2005) and older adults may be less inclined to use them (Levine, Lipsitz, & Linder, 2016). To break this so-called “law of attrition,” some mHealth studies involve older adults in the design process (Davidson & Jensen, 2013) or perform usability testing with older adults (e.g., Hong et al., 2014). However, there are age-related challenges to performing usability testing with older adults, such as caregiver interference, obstacles to using think-aloud, or location selection (Dickinson, Arnott, & Prior, 2007; Sonderegger, Schmutz, & Sauer, 2016). Further, while age may be an important consideration for mobile product testing, few studies have been published in this area (Franz, Munteanu, Neves, & Baecker, 2015; Grindrod, Li, & Gates, 2014), and fewer still concerning mHealth systems for seniors.

With respect to mHealth systems for geriatric heart failure, studies have reported user perceptions such as perceived learnability (Zan et al., 2015), but have stopped short of full usability evaluations to produce appropriate design principles.

Evaluating a Heart Failure Self-Management System

Based on literature and formative research with older adults with heart failure (Srinivas, Cornet, & Holden, 2016), we designed Engage, an mHealth system to be used by patients or informal caregivers for a 30-day period, during which their knowledge and motivation might improve, while lasting self-management routines are formed. Engage supports the setting and logging of self-management goals, recording and tracking of self-management data such as vitals and symptoms, and learning tips about heart failure self-management. The system

was also designed to securely store and communicate data between patients and designated individuals (e.g., clinicians).

In this paper, we report on a series of evaluation methods used to study the usability of Engage with a total of 15 older adults with heart failure and older adult informal caregivers.

METHODS

We tested Engage in two usability studies: one task-based (Study 1, $n=5$) and one scenario-based (Study 2, $n=10$).

Description of Engage, the Tested mHealth System

Engage was envisioned as a 30-day intervention for older adults with heart failure. It was originally designed for daily use on a mobile device (tablet or smartphone), with two daily sessions (morning, evening). In the morning, users received a list of actions to accomplish for the session, such as entering vitals, setting the day's self-management action plan, and reading tips about heart failure self-management. Some actions created follow-up actions for the evening session, for example, a check-up on whether the user achieved their action plan (Figure 1). Although the application was designed to support secure communication of these actions with clinicians, the usability tests focused on patients' independent product use.

Engage used an incentive-reward system with virtual coins earned for performing actions. The coin system was designed for flexibility, so different rewards (money, badges, coupons, etc.) could be redeemed based on the number of coins earned.

Figure 1 shows two screenshots of the Engage application. The left screen (Study 1) is titled 'Activity' and asks 'Did you do your activity goal today?'. It features a progress bar and a 'SUBMIT' button. The right screen (Study 2) is titled 'Physical Activity' and asks 'Today's plan was: 30min walk. Report what you actually did.' It features a progress bar and a 'SUBMIT' button.

Figure 1: Action plan report screens tested in Study 1 (L) and 2 (R)

Setting and Participants

Participants were English-speaking patients diagnosed with heart failure aged ≥ 60 years or an informal caregiver of such a person. They were recruited from two outpatient cardiology clinics, one urban and one suburban, of a Midwest academic health system. Each formally consented to the study and received a \$40 Visa gift card.

Study 1 had five patient participants (A1 to A5), three male, two female with a mean age of 61.2 ($SD = 4.97$). Study 2 had ten participants (B6 to B15), 8 patients and 2 informal caregivers, with a mean age of 70.8 ($SD = 7.96$). We collected additional demographics for Study 2: 8 participants were Caucasian, 2 were Black/African American; 5 participants completed at least some college; and 6 participants lived in households with annual income $\leq \$50,000$.

Procedure

In both Study 1 and 2, participants tested an interactive prototype of Engage created in Axure V7 on a 7-inch Asus tablet running the Android Operating System. During the tests, participants were instructed to think aloud concurrently with product use while being audio- and video-recorded.

At the start of each test, a researcher administered a structured interview assessing daily self-management routines and familiarity with technology. After the test, participants completed the System Usability Scale (SUS), the NASA Task Load Index (TLX) (Study 2 only), and a short debrief interview.

In Study 1 (Task-based), participants completed 8 tasks, each testing a component of Engage, e.g., action plan report (Figure 1). After each task, the researcher asked specific questions to test task comprehension and elicited feedback from participants, such as "how was that process for you?"

In Study 2 (Scenario-based), participants were presented a scenario of a fictitious sex-matched character (Jane or John). Participants were instructed to use Engage as if they were that person, based on information provided about the character's daily behaviors. After each simulated day in the life of the fictitious character, participants were queried about their understanding of what they had just done and general feedback.

The prototype was slightly redesigned between studies.

Instruments

SUS. The 10-item structured SUS was administered to assess overall product usability. The scale produces a score ranging from 0 to 100, with 68 being minimum acceptable usability (Brooke, 1996).

NASA TLX. The 6-item NASA TLX (Hart & Staveland, 1988) was administered to assess cognitive load associated with product use. We used the assessment of mental, physical, and temporal demand, performance, effort and frustration, and calculated an overall unweighted NASA TLX score.

RESULTS

Table 1 reports participants' technology use and ownership. Participants also differed in whether they recorded their daily weights (47%) and blood pressures (53%).

	Study 1 ($n=5$)	Study 2 ($n=10$)
Prior Tablet Use	3 (60%)	6 (60%)
Computer Ownership	4 (80%)	9 (90%)
Smartphone Ownership	2 (40%)	7 (70%)

Table 1. Participants' experiences with technology.

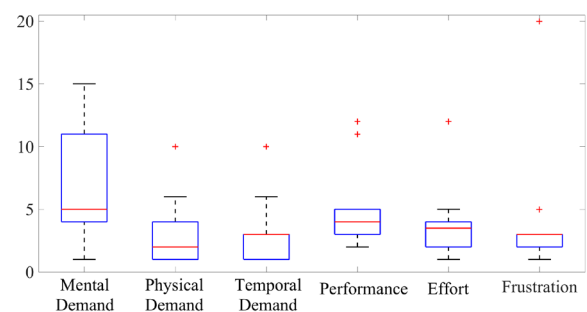


Figure 2: NASA-TLX component scores for Study 2 (lower is better).

Overall Findings from Standardized Evaluation Scales

Overall usability perceptions as assessed by SUS improved from Study 1 to Study 2 from a mean of 66.3 (SD=6.6) to 74.9 (SD=27.5). Removing an outlying score of 5 resulted in a mean Study 2 SUS score of 82.6 (13.2). Mean component NASA-TLX scores from Study 2 were highest for mental demand but generally low for other components (Figure 2).

Two participants' (A3 and B9) SUS scores rated the prototype as unnecessarily complex and very difficult to use. Among the two, B9 consistently gave low ratings for each of the SUS questions and reported high frustration on the TLX.

Findings from the Task-based Usability Test (Study 1)

Task Comprehension. For the seven questions assessing participants' understanding of the system, the average number of correct answers was 4.4 (63% accuracy, SD=2.1, range 2-7).

Understanding content. When asked to identify the content for one of the tasks (action planning), four (80%) participants could do so. However, some of the content was difficult to understand, such as selecting between the range-based options "less than 2000mg", "about 2000mg", "more than 2000mg", and "way above 2000mg". Two participants were unable to match values like 1750mg to the appropriate category. Even those who chose the right category were still confused, e.g.: "so what happened to 1750? I can't see the number... I picked less than 2000 because that is the only option I saw." (A4) Most participants (80%) could correctly tell the number of coins that they earned for performing an action in Engage. Three (60%) participants could consistently track the number of coins earned for their actions.

Difficulty of Use. For the first six tasks, participants were asked if the prototype was difficult to use. For five tasks, the majority ($\geq 60\%$) of participants reported that it was not difficult. However, 60% reported difficulty with the more complex task of setting and checking an activity plan, especially finding the edit function (accessible by tapping a pencil icon). The task with the most "not difficult" ratings (100%) was reading a tip about self-management.

Software Performance. One participant reported that the prototype was too slow (the prototype frequently slowed down and occasionally crashed during tests).

Findings from the Scenario-based Usability Test (Study 2)

Understanding the System's Purpose. 90% of all Study 2 participants could correctly state the purpose of Engage after completing the scenarios.

Willingness to Use the System. Most (70%) participants did not express any concern about using Engage for 30 days, the intended product use duration. However, participants did state concerns about adding a device on top of what they already have, getting sidetracked, and thus not using it every day.

Ease of Use. Initially, when asked how they felt about using Engage for their first part of the scenario, 80% of the participants reported that it was easy to use or not difficult to use. Throughout the subsequent parts of the scenario, 80% found the prototype easy to use ("If I could do it, anyone could", B11), and 60% reported that it became easier to use the prototype as they grew accustomed to it.

Perceptions about Gamified Elements. Study 2 participants were asked to reflect on the coin incentive system used in Engage, and reported mixed feelings. Half the participants felt they would be interested in earning coins only if the rewards were sufficiently motivating, e.g., real-world benefits, such as discounts on prescription drugs. The other half was attracted by simply earning the coins, independent of tangible rewards: "coins make you feel better about yourself" (B7) and "I'm a competitive person, so coins are an incentive to do something" (B13). Interestingly, the three participants least interested in the coins were the same ones who failed to accurately track the number of coins earned during the scenario: "I noticed the coins but I didn't really get the purpose. Like a pat on your shoulder?" (B15), and "The intrinsic value of the application is high enough to need not earn coins" (B14).

Perceptions of the Scenario-based Testing Process. Participants reflected on the data they were inputting during and after the scenario-based testing. Some participants reported difficulty working through the scenarios, indicating the product's potential flaws in accommodating real-world self-management. For example, four individuals had trouble mapping actual data to the options provided by the prototype; "it could be hard to estimate the [sodium intake] for the day" (B11). Similarly, the interpretation of "fluid", a word we used in the prototype, varied between participants; "you are saying you are going to drink 64oz of fluid, now the doctor [said] fluid doesn't mean tea or pop, it means water." (B11) and "As long as her coke was a diet coke, she's okay." (B8). These two individuals had technically incorrect mental models of heart failure and therefore struggled with the system's clinical assumptions (e.g., that all fluids count toward fluid restriction). Of course, some participants did have accurate mental models and used the system as intended: "Jane should've monitored what she was drinking during the day to not go over. She's gonna take more fluid, especially with Coke" (B15).

Conformity to the Scenario. Participants often identified with the persona from the scenario and were able to roleplay the use of Engage over the simulated 3-day period of use. Some participants empathized with the fictitious character and acted on their behalf: "she still walked, good for Jane, because it's hard to be outside that long, especially when it's that humid. So kudos for her" (B15). In contrast, some participants strayed from the scripted scenario, even making up reasons why the character might behave differently. For example, when setting the sodium intake plan for the day, one participant's thinkaloud report went like this: "she has to set her sodium intake level, but she knows that she's gonna go over. She doesn't want to do more than 2000. So she's gonna do less." (B8)

Resistance to using Engage. Two participants (B11 and B9) began the test expressing low self-efficacy or motivation to use Engage. One declared, "I don't think realistically that I would do this. If you give that to me to take home, I wouldn't do it. Even if the doctor gave it to me, I probably wouldn't be able to do it" (B9). The initial apprehension that the other participant with initial low motivation (B11) had eventually dissipated; this participant stated in his post-test answers that Engage would benefit himself as well as his doctor, by saving him some trips to the hospital and enabling his doctor to have his information instantly. B9 was however not convinced of the benefits he

could reap from Engage, but recognized that “there are people who like this kind of stuff... and got the time. So for these people it might be great.”

Participant Experiences with the Testing Procedures. Participants spontaneously mentioned difficulty performing the concurrent thinkaloud technique and explaining their thoughts, rather than simply describing their actions, e.g.: “*Talking to yourself is hard!*” (B11). One participant questioned the premise of using a fictitious scenario as the basis for the usability test: “*Well, if you consider a person’s time, it seems like a joke. You keep on telling me the answers. If I was putting my real information, it would be more meaningful for me and you.*” (B9). This individual also reported the lowest usability ratings and highest frustration with Engage.

DISCUSSION

mHealth System Usability Issues

The two studies resulted in different SUS scores. This may have been due to design modifications between the two studies, the use of different evaluation methods (tasks vs. scenarios), or sampling issues. It is also possible that the SUS is not a reliable or valid instrument for older adults across all education levels. The wording of the scale may have been problematic for certain people, as we have found in our previous experiences using SUS with older adults (Holden et al., 2016).

To assure product usability for older adult users, their age-related physical, cognitive, and attitudinal characteristics must be considered during design. Based on existing best practices for design for aging and disability (e.g., Fisk et al., 2009), we implemented larger text and controls than usual recommended by tablet application design guidelines and conventions (e.g., material design for the Android Operating System, available at material.io). While this design choice inconveniently reduced the amount of information that could be fit on each screen, it positively drove the realization of a simpler interface.

Nevertheless, participants desired an even simpler user interface with fewer steps and choices. Their comments indicated that they valued their time and disapproved of inefficiency, as predicted by socio-emotional theories of aging (Carstensen, Fung, & Charles, 2003) and models of technology acceptance by older adults (Chen & Chan, 2014). Test findings concerning efficiency resulted in subsequent redesigns, namely, reducing the number of steps to navigate the system and offering fewer options from which to select; e.g., the removal of the pencil-shaped “edit” button. The trade-off of these design decisions, however, was the reduced customizability and specificity of the system’s functionality. This raises the potential need to consider how design heuristics such as Nielsen’s “user control and freedom” apply to older adults who may value efficiency over choice (Nielsen & Molich, 1990).

Usability Testing with Older Adults as Participants

The results produced by SUS and NASA-TLX instruments contributed to the overall assessment of usability but did not “tell the whole story.” For example, the Study 2 participant reporting extremely low usability ratings and high frustration was more than a statistical outlier. He was visibly upset with the design of both the study and the system and questioned the

entire premise of a technology-based intervention, preferring being accountable to a human, not a machine. At the same time, he was able to navigate the system and perform the functions required to operate it with relative ease. This was not evident from the purely quantitative SUS and NASA TLX ratings, as older adults tend to have a different perception of their use of the system than their actual performance (Sonderegger et al., 2016). This case also highlights the need to consider affective design and individual differences (Khalid, 2006), in addition to technical usability and performance requirements.

In Study 1, by having participants complete multiple, smaller tasks, we were able to rapidly identify major design flaws, such as hiding options within menus. However, by deconstructing the test into discrete, vaguely related tasks, we were unable to test the way in which a person might actually use the system over a period of time. This led to a scenario-based, qualitatively driven approach in Study 2, allowing us to test users’ understanding of the purpose of using Engage for a full 30 days and testing scenarios in which data from earlier sessions affected later use behaviors. Most participants’ think-aloud reports were poor, as previous research on using the ‘think aloud’ approach with older adults has pointed out (Dickinson et al., 2007). However, the more frequent researcher-participant exchanges in Study 2 encouraged participants to describe their interaction with Engage with greater precision, counteracting the difficulties that older adults have with the ‘think aloud’ approach. Thus, Study 2 helped develop fuller accounts of how the system might be used, what issues might arise over time, and how the system did or did not accommodate the lived reality of illness self-management (Valdez, Holden, Novak, & Veinot, 2015). This is necessary to design mHealth applications meant to achieve long-term behavior change through extended use (Abedtash & Holden, 2017; Faiola & Holden, 2017).

All in all, the qualitative methods used with older adults yielded more valuable results than the two scales could provide.

Logistics of Conducting Usability Testing with Older Adults

We encountered remarkable challenges proper to conducting usability testing with older adults that could have played a role in the execution of the testing sessions.

Over a third of Study 2 participants were accompanied by one or two informal caregivers, who in most cases provided transportation to the testing session. These caregivers had the option to stay in the same room during the testing session; those who accepted provided valuable additional data, by correcting or completing participants’ statements about the self-management of their heart failure and the potential usefulness of Engage during the preliminary and post-study interviews.

We scheduled some testing sessions directly at the clinic to facilitate participants’ attendance, as some wanted to schedule the testing session back-to-back with their clinic appointment, and others preferred the familiarity of the clinic setting. This required extra flexibility of personnel schedule and transportation of hardware—cameras, tripods, microphones, documents—for testing sessions. Small-sized remote testing locations (such as patient examination rooms) conflicted with the hardware setup required for video recording, complicated the accommodation of participants with breathing devices or

wheelchairs, and in some cases prevented caregivers from staying in the room during the testing session.

Other recent studies have reported additional challenges in implementing health-related human factors research in community settings (Holden, McDougald Scott, Hoonakker, Hundt, & Carayon, 2015; Valdez & Holden, 2016). However, there are few resources for specifically considering the issues related to usability testing with older adults (Fisk et al., 2009; Sonderegger et al., 2016).

CONCLUSION

We generated a set of usability findings and redesign guidelines by triangulating the complementary results from task-based tests, scenario-based evaluation, and quantitative instruments. Beyond these, we identified multiple unanswered questions and future research direction regarding the process of usability testing with older adults.

ACKNOWLEDGMENTS

This study was sponsored by grant number K01AG044439 from the National Institute on Aging (NIA) of the US National Institutes of Health (NIH).

REFERENCES

- Abedtash, H., & Holden, R. J. (2017). Systematic review of the effectiveness of health-related behavioral interventions using portable activity sensing devices (PASDs). *Journal of the American Medical Informatics Association*, doi.org/10.1093/jamia/ocx1006.
- Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, & I. Lyall (Eds.), *Usability evaluation in industry* (pp. 189-194). London: Taylor & Francis.
- Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and emotion*, 27(2), 103-123.
- Chaudhry, S. I., McAvay, G., Chen, S., Whitson, H., Newman, A. B., Krumholz, H. M., & Gill, T. M. (2013). Risk Factors for Hospital Admission Among Older Persons With Newly Diagnosed Heart Failure: Findings From the Cardiovascular Health Study. *Journal of the American College of Cardiology*, 61(6), 635-642.
- Chen, K., & Chan, A. H. (2014). Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). *Ergonomics*, 57(5), 635-652. doi:10.1080/00140139.2014.895855
- Davidson, J. L., & Jensen, C. (2013). What health topics older adults want to track: a participatory design study. *Proceedings of ASSETS '13*.
- Dickinson, A., Arnott, J., & Prior, S. (2007). Methods for Human-Computer Interaction Research with Older People. *Behav. Inf. Technol.*, 26(4), 343-352.
- Eysenbach, G. (2005). The Law of Attrition. *J Med Internet Res*, 7(1), e11.
- Faiola, A., & Holden, R. J. (2017). Consumer Health Informatics: Empowering Healthy-Living-Seekers Through mHealth. *PROG CARDIOVASC DIS*, 59(5), 479-486.
- Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). *Designing for Older Adults: Principles and Creative Human Factors Approaches* (2nd ed.). Boca Raton, FL: CRC Press.
- Franz, R. L., Munteanu, C., Neves, B. B., & Baecker, R. (2015). *Time to Retire Old Methodologies? Reflecting on Conducting Usability Evaluations with Older Adults*. Paper presented at the Mobile HCI '15 Adjunct, Copenhagen, Denmark.
- Granger, B. B., Sandelowski, M., Tahshjain, H., Swedberg, K., & Ekman, I. (2009). A Qualitative Descriptive Study of the Work of Adherence to a Chronic Heart Failure Regimen: Patient and Physician Perspectives. *J Cardiovasc Nurs*, 24(4), 308-315.
- Grindrod, K. A., Li, M., & Gates, A. (2014). Evaluating User Perceptions of Mobile Medication Management Applications With Older Adults: A Usability Study. *JMIR mHealth and uHealth*, 2(1), e11.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In A. H. Peter & M. Najmedin (Eds.), *Advances in Psychology* (Vol. Volume 52, pp. 139-183): North-Holland.
- Holden, R. J., Bodke, K., Tambe, R., Comer, R. S., Clark, D. O., & Boustani, M. (2016). Rapid Translational Field Research Approach for eHealth R&D. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, 5(1), 25-27.
- Holden, R. J., McDougald Scott, A. M., Hoonakker, P. L. T., Hundt, A. S., & Carayon, P. (2015). Data collection challenges in community settings: Insights from two field studies of patients with chronic disease. *Quality of Life Research*, 24(5), 1043-1055.
- Holden, R. J., Schubert, C. C., Eiland, E. C., Storrow, A. B., Miller, K. F., & Collins, S. P. (2015). Self-care barriers reported by emergency department patients with acute heart failure: A sociotechnical systems-based approach. *Annals of Emergency Medicine*, 66(1), 1-12.
- Hong, Y., Goldberg, D., Dahlke, D. V., Ory, M. G., Cargill, J. S., Coughlin, R., . . . Peres, S. C. (2014). Testing Usability and Acceptability of a Web Application to Promote Physical Activity (iCanFit) Among Older Adults. *JMIR Human Factors*, 1(1), e2.
- Khalid, H. M. (2006). Embracing diversity in user needs for affective design. *Applied ergonomics*, 37(4), 409-418.
- Lainscak, M., Blue, L., Clark, A. L., Dahlström, U., Dickstein, K., Ekman, I., . . . Jaarsma, T. (2011). Self-care management of heart failure: practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. *European Journal of Heart Failure*, 13, 115-126.
- Levine, D. M., Lipsitz, S. R., & Linder, J. A. (2016). Trends in seniors' use of digital health technology in the united states, 2011-2014. *JAMA*, 316(5), 538-540. doi:10.1001/jama.2016.9124
- Mickelson, R. S., & Holden, R. J. (2013). Assessing the distributed nature of home-based heart failure medication management in older adults. *Proceedings of the Human Factors and Ergonomics Society*, 57(1), 753-757.
- Mickelson, R. S., & Holden, R. J. (2017). Medication adherence: staying within the boundaries of safety. *Ergonomics*, 1-22.
- Mickelson, R. S., Willis, M., & Holden, R. J. (2015). Medication-related cognitive artifacts used by older adults with heart failure. *Health Policy Techn*, 4, 387-398.
- Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Riegel, B., Dickson, V. V., & Topaz, M. (2013). Qualitative analysis of naturalistic decision making in adults with chronic heart failure. *Nursing Research*, 62, 91-98.
- Roger, V. L. (2013). Epidemiology of Heart Failure. *Circulation research*, 113(6), 646-659. doi:10.1161/CIRCRESAHA.113.300268
- Sonderegger, A., Schmutz, S., & Sauer, J. (2016). The influence of age in usability testing. *Applied Ergonomics*, 52, 291-300.
- Srinivas, P., Cornet, V., & Holden, R. (2016). Human Factors Analysis, Design, and Evaluation of Engage, a Consumer Health IT Application for Geriatric Heart Failure Self-Care. *Int J Hum-Comput Int*, 1-15. doi:10.1080/10447318.2016.1265784
- Valdez, R. S., & Holden, R. J. (2016). Health care human factors/ergonomics fieldwork in home and community settings. *Ergonomics in Design*, 24, 44-49.
- Valdez, R. S., Holden, R. J., Novak, L. L., & Veinot, T. C. (2015). Transforming consumer health informatics through a patient work framework: Connecting patients to context. *Journal of the American Medical Informatics Association*, 22(1), 2-10.
- van der Wal, M. H. L., & Jaarsma, T. (2008). Adherence in heart failure in the elderly: Problem and possible solutions. *Int J Cardiol*, 125, 203-208.
- van der Wal, M. H. L., Jaarsma, T., Moser, D. K., Veeger, N. J. G. M., van Gilst, W. H., & van Veldhuisen, D. J. (2006). Compliance in heart failure patients: The importance of knowledge and beliefs. *Eur Heart J*, 27, 434-440.
- van der Wal, M. H. L., Jaarsma, T., & van Veldhuisen, D. J. (2005). Non-compliance in patients with heart failure; how can we manage it? *Eur J Heart Fail*, 7, 5-17.
- Wu, J. R., Moser, D. K., Chung, M. L., & Lennie, T. A. (2008). Predictors of medication adherence using a multidimensional adherence model in patients with heart failure. *J Card Fail*, 14(7), 603-614.
- Zan, S., Agboola, S., Moore, S. A., Parks, K. A., Kvedar, J. C., & Jethwani, K. (2015). Patient Engagement With a Mobile Web-Based Telemonitoring System for Heart Failure Self-Management: A Pilot Study. *JMIR mHealth and uHealth*, 3(2), e33.